Final Report for Period: 09/2007 - 08/2008
Submitted on: 10/03/2008
Principal Investigator: Zha, Hongyuan .
Award ID: 0701796
Organization: GA Tech Res Corp - GIT
Submitted By:

Title:
Matrix Algorithms for Data Clustering and Nonlinear Dimension Reduction

Project Participants

Senior Personnel

Name: Zha, Hongyuan
Worked for more than 160 Hours: Yes
Contribution to Project:

Post-doc

Graduate Student

Name: Zhang, Ming
Worked for more than 160 Hours: Yes
Contribution to Project:
Ming Zhang carried out numerical experiments for the project

Undergraduate Student

Technician, Programmer

Other Participant

Research Experience for Undergraduates

Organizational Partners

Other Collaborators or Contacts

Activities and Findings

Research and Education Activities:
Major research and education activities:
Our research focuses on simultaneous clustering algorithms
and their applications in bioinformatics and information retrieval.
In particular, we proposed methods for learning ranking functions
by exploring the difference in query distributions (ref. 1). We also proposed robust algorithms for PCA and clustering by incorporating rotational invariance (ref. 2). We applied simultaneous clustering methods for document segmentation (ref. 3, ref. 4).

During this period, I moved to College of Computing at Georgia Tech where I developed a new course on 'Web search and text mining'

graduate course. I also further investigate applications of clustering and dimension reduction methods for citation analysis and document recommendations (ref. 5 and ref. 6).

1) Hongyuan Zha, Z. Zheng, H. Fu and G. Sun, Incorporating Query
Difference for Learning Retrieval Functions in Information Retrieval,
Proceedings of CIKM, 2006.

2) C. Ding, D. Zhou, X. He and Hongyuan Zha, R1-PCA: Rotational Invariant L1-norm Principal Component Analysis for Robust Subspace Factorization, Proceedings of ICML, 2006.

3) B. Sun, D. Zhou, H. Zha, J. Yen, Multi-Task Text Segmentation and
Alignment Based on Weighted Mutual Information. Proceedings of CIKM, 2006.

4) B. Sun, P. Mitra, L. Giles, Hongyuan Zha and J. Yen. Topic Segmentation with Shared Topic Detection and Alignment of Multiple Documents. Proceedings of SIGIR, 2007.

5) D. Zhou, I. Councill, Hongyuan Zha, C. L. Giles. Discovering Temporal Communities from Social Network Documents, ICDM 2007.

6) D. Zhou, S. Zhu, K. Yu, X. Song, B. Tseng, Hongyuan Zha, C. L. Giles. Learning Multiple Graphs for Document Recommendations. WWW, 2008.

Findings:

Implicitly clustering queries into different types can significantly
afftect the performance of the learned ranking. But at which granularity
to cluster the queries is still an open problem that deserves more in-depth research.

Training and Development:

Jiang Bian is a graduate student supported by this grant. Hongyuan Zha also developed a graduate course 'Web search and text mining' at Georgia Tech.

Outreach Activities:

Journal Publications

X. Ji and H. Zha., "Robust Sensor Localization Algorithm in Wireless Ad Hoc Sensor Networks.", Proceedings of the Twelfth International Conference on Computer Communications and Networks, p. 527, vol. , (2003). Published,

X. Ji and H. Zha., "Multidimensional Scaling Based Sensor Positioning Algorithms in Wireless Ad Hoc Sensor Networks.", Proceedings of the First ACM Conference on Embedded Networked Sensor Systems (SenSys'03), p. 328, vol. , (2003). Published,

H. Han, E. Manavoglu, H. Zha, K. Tsioutsiouliklis, L. Giles, "Rule-based Word Clustering for Document Metadata Extraction", Proceedings of the 20th Annual ACM Symposium on Applied Computing Special
Track on Information Access and Retrieval, p. 1058, vol. , (2005). Published,

Y. Zhang, X. Ji, C. H. Chu, and H. Zha, "Correlating Summarization of Multi-source News with K-Way Graph Biclustering", ACM SIGKDD explorations, p. , vol. , (2005). Accepted,

H. Han, H. Zha, L. Giles, "Name Disambiguation in Author Citations using a K-way Spectral Clustering Method", Proceedings of ACM/IEEE Joint Conference on Digital Libraries, p. , vol. , (2005). Accepted,

Y. Zhang, H. Zha, and C. Chu, "Inferring Interacting Domains: Challenges and Solutions", Annual International conference on Intelligent Systems for Molecular Biology, p. , vol. , (2005). Submitted,

D. Zhou, J. Li, H. Zha, "A Mallows Distance Based Metric for Comparing Clusterings", Proceedings of ICML, p. , vol. , (2005). Submitted,

Hongyuan Zha, Z. Zheng, H. Fu and G. Sun, "Incorporating Query Difference for Learning Retrieval Functions in Information Retrieval", Proceedings of CIKM, p. , vol. , (2006). Published,

C. Ding, D. Zhou, X. He and Hongyuan Zha, "R1-PCA: Rotational Invariant L1-norm Principal Component Analysis for Robust Subspace Factorization", Proceedings of ICML, p. , vol. , (2006). Published,

B. Sun, D. Zhou, Hongyuan Zha, J. Yen, "Multi-Task Text Segmentation and Alignment Based on Weighted Mutual Information", Proceedings of CIKM, p. , vol. , (2006). Published,

B. Sun, P. Mitra, L. Giles, Hongyuan Zha and J. Yen, "Topic Segmentation with Shared Topic Detection and Alignment of Multiple Documents", Proceedings of SIGIR, p. , vol. , (2007). Published,

Books or Other One-time Publications

D. Zhou, I. Councill, Hongyuan Zha, C. L. Giles, "Discovering Temporal Communities from Social Network Documents", (2007). Conference Proceedings, Published
Editor(s): ICDM
Collection: IEEE International Conference on Data Mining
Bibliography: IEEE Press

D. Zhou, S. Zhu, K. Yu, X. Song, B. Tseng, Hongyuan Zha, C. L. Giles, " Learning Multiple Graphs for Document Recommendations", (2008). Conference Proceedings, Published
Editor(s): World Wide Web
Collection: Proceedings of World Wide Web Conference
Bibliography: World Wide Web

Web/Internet Site

URL(s):
http://www.cse.psu.edu/~zha/papers.html
Description:

Other Specific Products

Contributions

Contributions within Discipline:

Implicitly clustering queries into different types can significantly afftect the performance of the learned ranking. But at which granularity to cluster the queries is still an open problem that deserves more in-depth research.

Contributions to Other Disciplines:
The applications of clustering to better organize citation information and improve retrieval relevance.

Contributions to Human Resource Development:
Jiang Bian is supported by the grant as a graduate research assistant.

Contributions to Resources for Research and Education:
Our research results are being incorporated into a graduate course on web search and text mining

Contributions Beyond Science and Engineering:

Categories for which nothing is reported:

Organizational Partners

Activities and Findings: Any Outreach Activities

Any Product

Contributions: To Any Beyond Science and Engineering